Readings

Matthew Kwan

March 10, 2014

Contents

1 Combinatorial estimates by the switching method [HM10]

1.1 Summary

Consider a finite set Ω of objects. A switching is a (nondeterministic) operation that transforms one object into another (or more generally, a switching is a relation $R \subseteq \Omega \times \Omega$). We can partition Ω into subsets $\{C(v)\}_{v \in V}$ and put a directed graph structure (possibly with loops) on the index set V: if an element in $C(v)$ can switch to an element in $C(w)$ then there is an arc between v and w.

Hopefully, for each v we can find a good lower bound $a(v)$ on the number of ways an $\omega \in C(v)$ can be switched, and a good upper bound $b(v)$ on the number of switchings an $\omega \in C(v)$ can be produced by. If we imagine that all switchings are performed at once, a and b give bounds on the inflow and outflow at each vertex in terms of the size of each $C(v)$. (So, $a(v)$ and $b(v)$ can more generally be bounds on the *average* number of switchings per element in $C(v)$).

This gives some information about the relative sizes of the classes $C(v)$. Given a set of vertices X, let $N(X)$ be the total amount of elements in all $C(x)$, where $x \in X$. The objective of the

paper is to bound $N(Y)/N(X)$ for given vertex sets X and Y.

In order to obtain bounds on $N(Y)/N(X)$, we write the constraints as a system of linear inequalities. Let $s'(v,w)$ be the (unknown) amount of switchings going from $C(v)$ to $C(w)$. For instance, for any vertex v, we have $\sum s'(vw) \geq a(v)N(v)$. A nonzero assignment of s and N satisfying the full set of inequalities is called a *feasible solution*, and a feasible solution which maximizes $N(Y)/N(X)$ is called an *optimal solution*.

The key result of the paper is to show that optimal solutions always take one of six standard forms. This is proved by a reduction of the inequalities to a linear program, and the fact that an optimal solution of a linear program always occurs at a vertex of the corresponding convex polytope. The analysis is amenable to a general bound α on the arcs of the digraph instead of the functions a, b (in that case we have $\alpha(v,w) = b(w)/a(v)$). Althoughα doesn't have a clear combinatorial interpretation, the paper suggests that different α may arise from richer information bounding the behaviour of the flow of switchings.

For a number of commonly-satisfied assumptions, the paper proves some alternative bounds for $N(Y)/N(X)$ that are slightly looser but easier to apply. In particular, a common use case is that we have some statistic on the objects in Ω and we believe that objects with a high value of that statistic make up a negligible fraction of all the objects in Ω. We would then partition the objects according to our statistic, and design a switching that tends to decrease the statistic, but by no more than some fixed amount. We would choose X to be the set of all vertices (partititions) and Y would be the set of all vertices (partitions) with a statistic value higher than some M.

1.2 Remarks

- I'm not sure why $i_{k-1} = \max\{M-(k-1)K, N+1\}$ is required for cases (a) and (b) of Lemma 3, instead of just $M-(k-1)K$ as in case (c).
- In Corollary 1, it isn't spelled out what assumptions X should satisfy, but it would seem A1 still has to hold.

- In Corollary 1, the somewhat inscrutable expression $k = \left\lceil \frac{M+\min\{0,K-\rho-1\}}{K} \right\rceil$ can be more easily seen to be $k = \left\lceil \frac{M-\lfloor \rho \rfloor}{K} \right\rceil$.

2 Asymptotic enumeration of sparse multigraphs with given degrees [GM13]

2.1 Summary

The paper introduces a generalized version of the switching method, where there are multiple different "colours" of switchings that operate on the same set of objects. By slightly loosening the bound, this generalized model can be reduced to an instance of the single-switching model.

The purpose of the generalized method is to make it tractable to generalize an earlier paper [MW91] which counted the number of graphs with a given degree sequence. The result was generalized to multigraphs.

2.2 Remarks

- The switchings are described in terms of *oriented* edges. That makes sense because M is the sum of the degrees and therefore twice the number of edges.
- When giving a lower bound for the number of switchings, there are formulas like $[\ell_1]_3 - O\big(k_{\max}\ell_1^2\big)$. The reasoning for the $O\big(k_{\max}\ell_1^2\big)$ term is that it bounds the number of triples where two vertices are loops and the second of those is adjacent to the final vertex. This is an upper bound for the number of triples of loops which contain an edge.

3 Random Graphs [JŁR00]

3.1 Remarks

3.1.1 Proposition 1.15

The "subsubsequence principle" means that taking n along any subsequence of $\mathbb{N}$, we can find a subsubsequence so that either M is bounded, or $N-M$ is bounded, or $M(N-M)/N \to \infty$. For, suppose none of these are true. Take a subsubsequence where M/N and $(N-M)/N$ both converge. They cannot both converge to zero; if for example $M/N \to L > 0$ then $M(N-M)/N = \Theta(N-M)$, contradiction.

In fact we can assume that if $M(N-M)/N \not\to \infty$ then M or $N-M$ is constant, which simplifies the analysis significantly. If M is constant, prove that $\mathbb{P}\big(|\Gamma_{M/N}| < M\big) \to p_M(M)/e^M$, where p_M is the first M terms of the taylor expansion of e^z. Also, I think the result of the central limit theorem is not enough for the estimate $\mathbb{P}\big(|\Gamma_{M/N}| < M\big) \to 1/2$, one needs to dig into the proof by convergence of characteristic function.

3.1.2 Theorem 5.4

The process can be viewed as either exploring a random graph or randomly generating a graph in an exotic edge order. Both are equivalent, but different viewpoints help understand different parts of the argument. The "kth step" means the kth saturation, not the kth generation. If the process is dead before step k, then X_k is not defined. The last paragraph on p109/first paragraph on p110 is confusing and contains some typos, here's a little reprashing of bits and pieces:

The event that after the first k steps (saturations) there are fewer than $(c-1)k/2$ unsaturated vertices remaining and the process is alive is

$$H = \left\{ \text{process is alive for step } k \text{and} \sum_{i=1}^{k} X_i - k < \frac{(c-1)k}{2} \right\}.$$

Now, at step i, if m vertices have been touched there are $n-m$ vertices that can be touched, each with probability p, so $X_i \sim \mathrm{Bi}(n-m,p)$. For event H, no more than $(c+1)k_+/2$ vertices can have been touched up to step k, so $X_i \geq X_i^-$ where $X_i^- \sim \mathrm{Bi}(n-k_+(c+1)/2,p)$. Then $\sum_{i=1}^k X_i^- \sim \mathrm{Bi}\big(k\big(n-n^{2/3}(c+1)\big)/2,p\big)$ and $\mathbb{E}\Big[\sum_{i=1}^k X_i^-\Big] > k-1$ for sufficiently large n.

The bound for there to be two giant components uses $e^Q \geq 1+Q$ for $Q=-c/n$. I think formal justification for the for the $o(1)$ term in $\rho_- + o(1)$ is quite complicated. Let $Y_n \sim \mathrm{Bi}(n,c/n)$ and $Y \sim \mathrm{Po}(c)$. Since $Y_n \xrightarrow{d} Y$, for all ε there is N so that $Y_n \leq (1+\varepsilon)Y$ for $n > N$. The pgf of Y is $G : x \mapsto \exp\big(c\big(x^{1+\varepsilon}-1\big)\big)$ so by the implicit function theorem, the solution for x of $G(x;\varepsilon)=x$ is continuous in ε and

$$\begin{aligned}&\mathbb{P}(\text{branching process with } X_i \sim \mathrm{Bi}(n,c/n) \text{ dies after more than } k_- \text{ steps})\\ &\leq \mathbb{P}(\text{branching process with } X_i \sim \mathrm{Po}(c) \text{ dies after more than } k_- \text{ steps}) + o(1)\end{aligned}$$

because $\mathbb{P}$(process dies at all) is (in the limit) the same for both cases. But since $k_- \to \infty$ the above probability is $o(1)$.

To choose an ordered pair of small vertices, first choose a small vertex arbitrarily (there are $n\rho(n,p)$ ways to do this on average), then either choose a small vertex in the same component (there are at most $k_- - 1$ ways to do this), or choose a small vertex in a different component (there are independently $n\rho(n-O(k_-),p)$ ways of doing this on average). Using a similar bounding argument as for $\rho(n,p)$ we have $\rho(n-O(k_-),p) \sim \rho(n,p)$.

3.1.3 Theorem 5.5

The reasoning behind the estimate $k!k^2$ is that each bad subgraph can be constructed by choosing a path then choosing two points on that path to glue the endpoints to. The first line of the displayed inequality is due to Markov's inequality, the second can be derived with $\big(\frac{2M}{n}\big)^{k+1} > \frac{n^{k+1}[M]_{k+1}}{[n^2/2-1]_{k+1}} > \frac{[n]_k[M]_{k+1}}{\left[\binom{n}{2}\right]_{k+1}}$ for $2 \leq k < n/2$.

3.1.4 Proof of Theorem 9.23

For the trace identity, prove by induction that $\left(A^k\right)_{ij}$ is the sum of all products $a_{i\alpha_1}a_{\alpha_1\alpha_2}\dots a_{\alpha_{k-2}\alpha_{k-1}}a_{\alpha_{k-1}j}$.

4 Fundamentals of Stein's Method [Ros11]

4.1 Summary

4.1.1 Overview

Stein's method is a technique for bounding the "distance" between distributions in some metric of the form $d(X,Z) = \sup_{h\in\mathcal{H}}|\mathbb{E}h(X) - \mathbb{E}h(Z)|$, where $\mathcal{H}$ is a family of functions. If $\mathcal{H} = \left\{1_{(-\infty,x]} : x \in \mathbb{R}\right\}$, then this measures the maximum distance between distribution functions (Kolmogorov metric). It's often possible to transfer bounds between different metrics, so it's worthwhile to define some auxilary metrics for which it's easier to apply Stein's method.

I think the motivations behind Stein's method are twofold. First, it gives a concrete bound of the kind not usually provided by asymptotic methods. Second, it can be used to give uniform convergence of probabilities, which is sometimes a necessary ingredient to asymptotic results where we allow things to vary (for example, proving something about $\mathbb{G}(n,p)$)

The idea of Stein's method is that there is some characterizing operator $\mathcal{A}$ of a random variable Z so that $\mathbb{E}\mathcal{A}f(X) = 0$ for all f in some manageable family $\mathcal{F}$, precisely when X has the distribution of Z. For example, if $Z \sim \mathcal{N}(0,1)$ then we can choose $\mathcal{A}$ as $f \mapsto (x \mapsto f'(x) - xf(x))$ and $\mathcal{F}$ as the set of bounded, continuously differentiable functions. If $Z \sim \text{Po}(\lambda)$ then we can choose $\mathcal{A}$ as $f \mapsto (k \mapsto \lambda f(k+1) - kf(k))$ and $\mathcal{F}$ as the set of all bounded functions with bounded differences.

For all $h \in \mathcal{H}$, we can try to choose f_h to solve (or approximately solve) the functional equation $\mathcal{A}f_h = h - \mathbb{E}hZ$ (and adjust/refine $\mathcal{F}$ to include all such f_h). Then, $d(X,Z) \leq \sup_{f\in\mathcal{F}}|\mathbb{E}\mathcal{A}f(X)|$. By the choice of $\mathcal{A}$, we hope that this bound is relatively sharp. Also, we also hope that bounding $|\mathbb{E}\mathcal{A}f(X)|$ is easier than directly bounding $|\mathbb{E}h(X) - \mathbb{E}h(Z)|$.

4.1.2 Size-Bias Coupling

For a nonnegative integrable random variable X, we say X^s has the size-bias distribution with respect to X if for all f so that $Xf(X)$ is integrable, we have $\mathbb{E}[Xf(X)] = \mathbb{E}X\mathbb{E}f(X^s)$. For example, if X is a Bernoulli random variable then we can choose $X^s = 1$. The size-bias distribution always exists, with $\mathrm{d}(\mathcal{L}X^s)(x) = \frac{x}{\mathbb{E}X}\,\mathrm{d}(\mathcal{L}X)(x)$; that is $(\mathcal{L}X^s)(A) = \mathbb{E}[X\mathbf{1}_A \circ X]/\mathbb{E}X$. The size-bias distribution can be seen as weighting (biasing) the probability of X taking some value x by the size of x.

The idea is that we can perturb a nonnegative random variable X to obtain X^s in such a way that $X - X^s$ has low variance. If X has mean μ and variance σ^2, then $\mathbb{E}[X^s - X] = \frac{1}{\mu}\mathbb{E}\big[X^2\big] - \mu = \frac{\sigma^2}{\mu}$. If X is approximately Poisson, then this is equal to 1, so if $X - X^s$ has low variance then $\mathbb{E}[Xf(X) - \lambda f(X+1)] = \lambda\mathbb{E}[f(X^s) - f(X+1)] \approx 0$

For the normal case, define $W = \frac{X-\mu}{\sigma}$. If $X^s - X$ has low variance then

$$\begin{aligned}
\mathbb{E}[Wf(W)] &= \frac{\mu}{\sigma}\mathbb{E}\left[f\left(\frac{X^s-\mu}{\sigma}\right) - f\left(\frac{X-\mu}{\sigma}\right)\right] \\
&\approx \frac{\mu}{\sigma}\mathbb{E}\left[f'(X)\left(\frac{X^s - X}{\sigma}\right)\right] \\
&\approx \frac{\mu}{\sigma^2}\mathbb{E}\big[f'(X)\big]\mathbb{E}[X^s - X] \\
&= \mathbb{E}\big[f'(X)\big].
\end{aligned}$$

Here's an often-useful description of the size-bias distribution of a sum of random variables: If $\mathbf{X} = (X_i)_{i=1}^n$ is a sequence of random variables, define $\mathbf{X}^{(i)}$ by $\mathrm{d}\big(\mathcal{L}\mathbf{X}^{(i)}\big)(\mathbf{x}) = \frac{x_i}{\mathbb{E}X_i}\,\mathrm{d}(\mathcal{L}\mathbf{X})(\mathbf{x})$, so that for any f, we have $\mathbb{E}\big[f\big(\mathbf{X}^{(i)}\big)\big] = \mathbb{E}[X_i f(\mathbf{X})]/\mathbb{E}[X_i]$. Let $X = \sum_{i=1}^n X_i$ and let I satisfy $\mathbb{P}(I = i) = \mathbb{E}[X_i]/\mathbb{E}[X]$. Then $\mathcal{L}X^s = \mathcal{L}\sum_{i=1}^n X_i^{(I)}$.

An interpretation of $\mathrm{d}\big(\mathcal{L}\mathbf{X}^{(i)}\big)(\mathbf{x}) = \frac{x_i}{\mathbb{E}X_i}\,\mathrm{d}(\mathcal{L}\mathbf{X})(\mathbf{x})$ for discrete random variables is that the ith component of $X^{(i)}$ has the size-bias distribution of X_i and given $X_i^{(i)} = x$, the other components of $\mathbf{X}^{(i)}$ have the distribution of the corresponding components of $\mathbf{X}$, conditioned on $X_i = x$.

In practice, we can often represent X as a sum of indicator variables (which have size-bias

distribution 1), each representing some local property of an underlying space of combinatorial objects. For example, X might be the number of isolated vertices in a random graph $G \in \mathbb{G}(n,p)$, so we would have an indicator for each vertex. For each G we choose a location at random and force the local property to hold there; this creates a new object G'. Provided we have made the local adjustment in the right way, we will have $\mathcal{L}X(\cdot') = \mathcal{L}X^s$ in such a way that X is close to X^s. For the isolated vertex example, we can create G' by deleting all the edges incident to a randomly chosen vertex.

Apparently, size-bias coupling is extensively used in [BHJ92], but not by name.

4.1.3 Exchangeable Pairs for Poisson Approximation

A pair of random variables (X, X') is exchangable if it has the same distribution as (X', X). If (X, X') are exchangeable, we have $\mathbb{E}[1\{X = X' + 1\}f(X)] = \mathbb{E}[1\{X' = X + 1\}f(X')]$ so $\mathbb{E}[\mathbb{P}(X' = X - 1|X)f(X)] = \mathbb{E}[\mathbb{P}(X' = X + 1|X)f(X + 1)]$.

So, if we have $\mathbb{P}(X' = X - 1|X) \approx cX$ and $\mathbb{P}(X' = X + 1|X) \approx c\lambda$ for some c then $\mathbb{E}[\lambda f(X) - Xf(X)] \approx 0$ and X is approximately Poisson.

To choose appropriate X', we want to perturb X in some "reversible" way, so that X can also be symmetrically viewed as a perturbation of X'. Sometimes X can be viewed as some statistic on a space Ω of combinatorial objects with some distribution π, so we can obtain X' with a reversible Markov chain on Ω that has stationary distribution π.

For example, if X is a sum of indicator variables $X_1, \ldots, X_n$, we could view X as the cardinality of a random subset of $\{1, \ldots, n\}$. For the transitions of a Markov chain, we could randomly select some $i \in \{1, \ldots, n\}$ and independently re-determine its membership. This sort of construction often naturally gives a suitable exchangeable pair. The event $\{X' = X - 1\}$ can be interpreted as a "death" and the event $\{X' = X + 1\}$ can be interpreted as an "immigration" in the Markov chain. The probability of a death is roughly proportional to the number alive, and the probability of an immigration is roughly proportional to the sum of the probabilities for each element to be alive in the stationary distribution, which is $\mathbb{E}[X] \approx \lambda$.

4.2 Remarks

- For the proof of proposition 1, the $C\varepsilon/2$ term comes from a $C \times \varepsilon$ triangular region of integration. The $d_W(W, Z)/\varepsilon$ term comes from the fact that $h_{x,\varepsilon}\varepsilon$ is 1-Lipschitz.
- For the proof of Lemma 3.4, as well as Cauchy-Schwarz Hölder's inequality $\mathbb{E}|X| \leq \left(\mathbb{E}X^2\right)^{1/2}$ is used.
- For the proof of Theorem 3.5, the fact $\sum_{i=1}^{n} \sum_{j \in N_i} X_i = \left(\sum_{i=1}^{n} X_i\right)^2$ is used.

5 Approximate Computation of Expectations [Ste86]

5.1 Summary

I also draw from [Ste92] and the first section of [CDM05], which provide similar explanations of the abstraction behind Stein's method for exchangeable pairs.

Stein's method was originally conceived by Stein as a method of approximately computing expectations. The common use of Stein's method to bound the distances between distributions follows naturally from this, but this paper gives insight as to why the methods for exchangeable pairs in [Ros11] work.

We redefine some notation from the discussion on [Ros11]. $\mathbb{E}_0$ is the expectation operator with respect to an approximating distribution (for example, Poisson). $T_0 : \mathcal{F} \mapsto \mathcal{H}$ is the characterizing operator of that distribution (for example, $f \mapsto (k \mapsto \lambda f(k+1) - kf(k))$). For T_0 to be characterizing means $\mathrm{im} T_0 = \ker \mathbb{E}_0$. We also want an operator U_0 which captures the idea in [Ros11] of solving for f_h: U_0 must satisfy $T_0 U_0 h = h - \mathbb{E}_0 h$.

From an exchangeable pair (X, X') we can construct an operator $T : F \mapsto \mathbb{E}^X[F(X, X')]$ from the set of antisymmetric functions $\mathcal{F}$. Under some trivially satisfied conditions, we have $\mathrm{im} T = \ker \mathbb{E}$. That is, T is a "characterizing operator" for $\mathbb{E}$. We then choose a connection operator $\alpha : \mathcal{F}_0 \to \mathcal{F}$.

With this data, Stein's lemma says that $\mathbb{E}_X = \mathbb{E}_0 + \mathbb{E} \circ R$ for some "remainder" operator R. If our choices were good then R is provably small. From here we can estimate distance metrics, or try to estimate specific expectations $\mathbb{E}_X h$.

5.2 Remarks

- We frequently use spaces of functions like $\mathcal{X}$ and $\mathcal{F}$. I think "space" means vector space

6 A way of Using Auxiliary Randomization[Ste92]

6.1 Remarks

- Page 14 says that a characterizing operator should satisfy $\ker \mathbb{E}_0 = \mathrm{im} T_0$, but i think this should really read $\mathcal{X}_0 \cap \ker \mathbb{E}_0 = \mathrm{im} T_0$ or $\mathbb{E}_0$ should be restricted somewhere (maybe it's implied, it's not very clear).

7 Exchangeable pairs and Poisson approximation[CDM05]

7.1 Remarks

- The "top row" and "bottom row" of the diagram seem to be mixed up.

8 Exchangeable Pairs, Switchings and Random Regular Graphs [Joh11]

8.1 Summary

A previous paper [MWW04] proved that small cycle counts were asymptotically Poisson under some "natural" conditions on the growth of the maximum cycle length, and conjectured that

this was a threshold; if the maximum cycle length was to grow faster the cycle counts would no longer be asymptotically Poisson. Johnson disproves the conjecture with Stein's method.

The use of switchings is quite different to Hasheminezhad and McKay. The switching operation (relation) induces a markov chain on the space of all regular graphs; by adding loops this chain can be adjusted so that the stationary distribution is uniform and the chain is reversible. The Markov chain then gives an exchangeable pair with which to apply Stein's method.

8.2 Remarks

- Proof of Theorem 7: I think every vertex of $\mathfrak{G}$ needs to have total *weight* d_0, not degree d_0.
- Unlike most (?) applications of Stein's method, Johnson is more interested in asymptotics than explicit quantitative bounds between distributions, so there are undetermined constants. I'm not sure whether it's feasible or useful to obtain these.

9 Stein's method and the rank distribution of random matrices over finite fields [FG12]

9.1 Summary

This paper uses the techniques of Poisson approximation in a more general context. If $X \sim \mathrm{Po}(\lambda)$ then $\lambda\mathbb{P}(X = k-1) = k\mathbb{P}(X = k)$. The methods are extended to the case where $a(k)\mathbb{P}(X = k-1) = b(k)\mathbb{P}(X = k)$. In particular, the limiting distribution of the rank of a random matrix over a finite field is known to have this form.

10 On Stein's method and perturbations [BČX07]

10.1 Summary

This paper uses Stein's method to compare distributions to "perturbations" of the Poisson and Normal distributions. Perturbations are defined in terms of "generators" of measures, which I probably need to learn about before tackling this paper properly. (The generator approach is discussed in [BC05]).

References

[BC05] Andrew D Barbour and Louis Hsiao Yun Chen, *An introduction to Stein's method*, Lecture notes series, Institute for Mathematical Sciences, National University of Singapore, vol. 4, Singapore University Press and World Scientific, 2005.

[BČX07] Andrew D Barbour, Vydas Čekanavičius, and Aihua Xia, *On Stein's method and perturbations*, arXiv preprint math/0702008 (2007).

[BHJ92] Andrew D Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, Clarendon Press Oxford, 1992.

[CDM05] Sourav Chatterjee, Persi Diaconis, and Elizabeth Meckes, *Exchangeable pairs and Poisson approximation*, Probab. Surv **2** (2005), 64–106.

[FG12] Jason Fulman and Larry Goldstein, *Stein's method and the rank distribution of random matrices over finite fields*, arXiv preprint arXiv:1211.0504 (2012).

[GM13] Catherine Greenhill and Brendan D McKay, *Asymptotic enumeration of sparse multigraphs with given degrees*, arXiv preprint arXiv:1303.4218 (2013).

[HM10] Mahdieh Hasheminezhad and Brendan D McKay, *Combinatorial estimates by the switching method*, Contemporary Mathematics (2010), 209–221.

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Cambridge Univ Press, 2000.

[Joh11] Tobias Johnson, *Exchangeable pairs, switchings, and random regular graphs*, arXiv preprint arXiv:1112.0704 (2011).

[MW91] Brendan D McKay and Nicholas C Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees* $o(n^{1/2})$, Combinatorica **11** (1991), no. 4, 369–382.

[MWW04] Brendan D McKay, Nicholas C Wormald, and Beata Wysocka, *Short cycles in random regular graphs*, Electron. J. Combin **11** (2004), no. 1, 1–12.

[Ros11] Nathan Ross, *Fundamentals of Stein's method*, Probab. Surv **8** (2011), 210–293.

[Ste86] Charles Stein, *Approximate computation of expectations*, IMS Lecture Notes – Monograph Series, no. 7, IMS, 1986.

[Ste92] ———, *A way of using auxiliary randomization*, Probability Theory (1992), 159–180.